

COMMUNICATION SYSTEM AND COMMUNICATION METHOD USING  
ANIMATION AND SERVER AS WELL AS TERMINAL DEVICE USED  
THEREFOR

This application is based on Japanese Patent Application Nos. 2000-176677 and 2000-176678 filed on June 13, 2000, the contents of which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] The present invention relates to a communication system using animation and a server as well as a terminal device used for the communication system. According to the present invention, a user accesses to a server from a client via a network so that the user can remotely perform a conversation while watching animation of an actual or fictional human or the like virtualized by using a computer. In addition, the user can converse while watching animation of a person to whom the user talks.

2. Description of the Prior Art

[0002] In recent years, a technique for communicating with an actual or fictional human, animal, doll or character that are virtualized by using a computer has been researched and developed.

[0003] For example, Japanese unexamined Patent Publication No. 11-212934 discloses a technique for having a creature that is raised in a virtual space perform a predetermined action by inputting a command via an input device such as a mouse or a keyboard. According to the

09873207-061201

technique, a user takes care of a virtual pet using a computer. Specifically, the user feeds the pet, lays the pet down, praises the pet, reproves the pet or plays with the pet in a similar way to taking care of a real pet by using a computer. The pet is raised by the user as described above and the user can experience how to raise pet with confirming growth of the virtual pet via images and voices output from a display or a speaker. It is also possible to remotely raise the pet via a network.

[0004] As a method for matching an output timing of voices of life with an output timing of images thereof, there is proposed a method disclosed in European Patent No. 0860811 in which the voices are synchronized with the images for output and a method disclosed in Japanese Unexamined Patent Publication No. 10-293860 in which the images are synchronized with the voices for output. Above method enables production of animation and output of the voices at the same time with the animation; therefore, the user can realistically recognize the output images and voices. As a method for producing animation based on actual film images, there is proposed an animation synthesis technique by way of recognition of actual film images (P.98-106, December 1998, NTT Technical Journal). According to the technique, a portrait is automatically made by a picture and expressions of different opening states of eyes and a mouth and expressions of various emotions are automatically made based on the portrait. Then, the portrait is synchronized with a voice so that portrait animation can be synthesized.

[0005] In the above-described technique disclosed in

Japanese Unexamined Patent Publication No. 11-212934, the user can remotely communicate with the virtual pet via the network. In the conventional technique, however, the virtual pet is controlled by commands from the user that are input via the input device so as to be displayed on the display; therefore, the user can communicate with the virtual pet only in limited patterns. For example, the technique does not allow conversation between the user and the virtual pet; therefore, realistic communication cannot be achieved by the technique.

[0006] The technique for producing the animation disclosed in European Patent No. 0860811 enables production of the animation including a motion of a person who is talking, for example. However, the user and the person cannot talk to each other, since the voices and the images are output uni-directionally from the person to the user.

[0007] A communication system such as a television telephone or a television conference system is actually utilized, in which a conversation can be performed with watching a partner's face by transmitting and receiving voices and images among a plurality of terminal devices.

[0008] However, since the images have a large amount of data, a communication line having large capacity for communication is required in order to transmit and receive the images. In the case of transmission and reception of images via a general telephone line, it is impossible to send and receive more than a few frames as an image per second and, therefore, it is impossible to display a satisfactorily animated image. In turn, the usage of a

high-speed private line enables display of animated images wherein a motion appears substantially natural, however, it has not been widely prevalent yet due to high communication cost.

[0009] In order to reduce communications traffic, there has been proposed a method in which images of a part of and whole parts of a face are previously produced at low resolution for registration in a database, and then the whole facial image is displayed on a screen of a receiver's terminal device at the start of a conversation and only a part of the facial image corresponding to a part in which expressions have changed is downloaded from the database to the terminal device so as to be displayed in Japanese Unexamined Patent Publication No. 10-200882.

[0010] Reduction in the communications traffic can be realized by using the above-described conventional method. However, it is difficult to express a natural motion such as person's expressions since the resolution of the images is low and a plurality of two-dimensional images is continuously combined so as to be displayed.

[0011] Additionally, since respective users performing a conversation by means of the communication system must understand a common language, it is impossible for users using different languages to utilize the communication system described above.

#### SUMMARY OF THE INVENTION

[0012] An object of the present invention is to provide a communication system, a server and a client for achieving a remote conversation with an actual or

000282360 20240726T094201Z

fictional human or the like virtualized by using a computer.

[0013] Another object of the present invention is to reduce communications traffic and to perform a conversation with watching animation in which a motion of a partner (user at the other end) is smooth and substantially natural.

[0014] Further object of the present invention is to realize a conversation with watching animation in which a partner's motion is smooth and substantially natural even in a conversation between users using different languages.

[0015] According to one aspect of the present invention, a communication system for performing a conversation with an actual or fictional human, animal, doll or character virtualized by using a computer comprises a client and a server, wherein the client includes an input portion for inputting a first message addressed from a user to the human, the animal, the doll or the character, a transmitting portion for transmitting the first message, a receiving portion for receiving a second message which is a message addressed from the human, the animal, the doll or the character to the user as a response to the first message and facial animation of the human, the animal, the doll or the character, an output portion for outputting the second message to the user and a display portion for displaying the facial animation; and the server includes a storing portion for storing facial image data of the human, the animal, the doll or the character, a receiving portion for receiving the first message, a first generating portion for generating the second message in response to

00278607-00284360

the reception of the first message, a second generating portion for generating motion control data for moving the facial image data in accordance with the second message, a third generating portion for generating the facial animation based on the motion control data and the facial image data and a transmitting portion for transmitting the second message and the facial animation.

[0016] According to another aspect of the present invention, a communication system for performing a conversation with watching a partner's animation (animation of a partner) comprises a host computer and a plurality of terminal devices, wherein each of the terminal devices includes a transmission and reception portion for transmitting and receiving a voice, a first receiving portion for receiving image data, a second receiving portion for receiving motion control data for moving the image data and a display portion for displaying animation generated by moving the image data based on the motion control data, and the host computer includes a receiving portion for receiving a voice, a translation portion for translating the received voice into another natural language, a first transmitting portion for transmitting the translated voice, a generating portion for generating the motion control data based on the translated voice and a second transmitting portion for transmitting the image data and the motion control data of one of the terminal devices in communication to another one of the terminal device in the communication.

[0017] Further objects and advantages of the invention can be more fully understood from the following drawings

and detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] Fig. 1 is a block diagram showing a whole structure of a communication system according to the present invention.

[0019] Fig. 2 shows a program stored in a client of a first embodiment.

[0020] Fig. 3 shows a program stored in a server of the first embodiment.

[0021] Fig. 4 shows a database provided in a magnetic disk unit of the server.

[0022] Fig. 5 shows an example of a person list.

[0023] Fig. 6 is a flowchart showing a process of a communication system of the first embodiment.

[0024] Fig. 7 is a flowchart showing a process for generating facial animation data and a second message.

[0025] Fig. 8 generally shows an example of facial image data.

[0026] Fig. 9 shows a program stored in a client of a second embodiment.

[0027] Fig. 10 shows a program stored in a server of the second embodiment.

[0028] Fig. 11 is a flowchart showing a process of a communication system of the second embodiment.

[0029] Fig. 12 is a flowchart showing a process for generating motion control data and a second message.

[0030] Fig. 13 is a block diagram showing databases stored in each magnetic disk unit of a client and a server according to a third embodiment.

[0031] Fig. 14 is a block diagram showing a whole structure of a communication system according to a fourth embodiment of the present invention.

[0032] Fig. 15 shows an example of a program and data stored in a terminal device.

[0033] Fig. 16 shows an example of a program and data stored in a host computer.

[0034] Fig. 17 is a flowchart showing a process of the communication system.

[0035] Fig. 18 is a flowchart showing a process of the terminal device.

[0036] Fig. 19 is a flowchart showing a process of the host computer.

[0037] Fig. 20 shows an example of a program and data stored in a terminal device of a fifth embodiment.

[0038] Fig. 21 shows an example of a program and data stored in a host computer.

[0039] Fig. 22 is a flowchart showing a process of a communication system.

[0040] Fig. 23 is a flowchart showing a process of a terminal device.

[0041] Fig. 24 is a flowchart showing a process of a host computer.

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0042] First, as a communication system, three embodiments will be described. In communication systems 1, 1B and 1C of the three embodiments, various persons may be virtualized by using a computer and may be displayed as animation. A user can select a person according to the

user's preference from the persons and perform a conversation with the selected person.

(First Embodiment)

[0043] Fig. 1 is a block diagram showing a whole structure of a communication system 1 according to a first embodiment of the present invention. Fig. 2 shows an example of a program stored in a magnetic disk unit 27 in a client 2. Fig. 3 is a diagram showing an example of a program stored in a magnetic disk unit 37 in a server 3. Fig. 4 shows an example of a database provided in the magnetic disk unit 37 in the server 3. Fig. 5 generally shows an example of a list LST of a person HMN.

[0044] As shown in Fig. 1, a communication system 1 comprises a client 2, a server 3, and a network 4.

[0045] The client 2 includes a processor 21, a display 22a, a speaker 22b, a mouse 23a, a keyboard 23b, a microphone 23c, a communication controller 24, a CD-ROM drive 25, a floppy disk drive 26 and a magnetic disk unit 27.

[0046] The processor 21 has a CPU 21a, a RAM 21b and a ROM 21c so as to execute a series of processes in the client.

[0047] The RAM 21b temporarily stores a program or data or the like, while the ROM 21c stores a program and set information of hardware of the client and the like. The CPU 21a executes the programs.

[0048] The display 22a displays animation of a face of a person HMN and outputs after-mentioned character data TXT2 in the form of display. The speaker 22b outputs after-mentioned voice data SND2 below as a voice. The

mouse 23a and the keyboard 23b are used for inputting a first message MG1 as a message addressed from the user to the person HMN, or for operating the client 2, or the like. The microphone 23c is used for inputting the first message MG1 in the form of the voice.

[0049] The communication controller 24 controls transmission and reception of the first message MG1, a second message MG2 which is a message addressed from the person HMN to the user, facial animation data FAD to be described below, and other data. The CD-ROM drive 25, the floppy disk drive 26 and the magnetic disk unit 27 store data and programs.

[0050] The server 3 includes a processor 31, a display 32, a mouse 33a, a keyboard 33b, a communication controller 34, a CD-ROM drive 35, a floppy disk drive 36 and a magnetic disk unit 37.

[0051] The processor 31 comprises a CPU 31a, a RAM 31b and a ROM 31c. The structure and the function of the processor 31 are the same as those of the above-described processor 21. The communication controller 34 controls transmission and reception of the first message MG1, the second message MG2, the facial animation data FAD and other data.

[0052] The network 4 comprises a public line, a private line, a LAN, a wireless line or the Internet. The client 2 and the server 3 are connected with each other via the network 4.

[0053] The first message MG1 includes voice data SND1 input from the microphone 23c or character data TXT1 input from the keyboard 33b. The second message MG2 includes

the voice data SND2 or the character data TXT2. The facial animation data FAD are information of facial animation comprising images indicating continuous motion of a face of a person HMN.

[0054] As shown in Fig. 2, the magnetic disk unit 27 in the client 2 stores an OS 2s as a basic program of the client 2, a client conversation program 2p as an application program of the client in the communication system 1, data 2d required therefor and the like. The client conversation program 2p serves to carry out a basic operation process 2bs and other processes. The basic operation process 2bs is a process for performing linkage with the OS 2s, operations relative to a selection of a person HMN and input of the first message MG1. The programs and data are loaded into the RAM 21b as required so as to be executed by the CPU 21a.

[0055] As shown in Fig. 3, the magnetic disk unit 37 in the server 3 stores an OS 3s as a basic program of the server 3, a server conversation program 3p as an application program of the server in the communication system 1, data 3d which are information required therefor and the like.

[0056] The server conversation program 3p comprises a basic operation process 3bs, a language recognition conversation engine EG1 and an animation engine EG2. The basic operation process 3bs is a process for performing linkage with the OS 3s. The basic operation process 3bs is also a process for supervising and controlling the language recognition conversation engine EG1 and the animation engine EG2.

[0057] The language recognition conversation engine EG1 is a system for performing a language recognition process 3gn and a conversation generating process 3ki and the system is known. The language recognition process 3gn is a process for analyzing the voice data SND1 to extract character data TXTa expressed by natural languages such as Japanese or English. The conversation generating process 3ki is a process for generating the voice data SND2 or the character data TXT2.

[0058] In order to produce the voice data SND2, voice data of an identical person or of a substitute person are previously obtained with respect to each of the person HMN. Voice synthesis is performed by the conversation generating process 3ki based on the obtained voice data.

[0059] The animation engine EG2 carries out a motion control process 3ds and an animation generating process 3an. Motion control data DSD are generated by the motion control process 3ds. The motion control data DSD are control information for controlling facial image data FGD of the person HMN in such a manner that the facial image data FGD of the person HMN move in accordance with a timing of output of the second message MG2 from the speaker 22b or the display 22a. The animation generating process 3an is a process for generating the facial animation data FAD based on the motion control data DSD and the facial image data FGD.

[0060] The programs are suitably loaded into the RAM 31b so as to be executed by the CPU 31a. Further, if required, the RAM 31b temporarily stores the first message MG1, the facial image data FGD, the second message MG2, the motion

control data DSD, the facial animation data FAD and the like all of which are used for these processes.

[0061] As shown in Fig. 4, the magnetic disk unit 37 is provided with a facial image database FDB, a person information database HDR and a conversation database KDR.

[0062] The facial image database FDB accumulates the facial image data FGD of persons HMN. The person information database HDB includes person information HMJ that is information of gender, character, age and the like of each of the persons HMN. The conversation database KDB accumulates sentence information BNJ and word information TNJ as grammar and words for generating sentences for conversation.

[0063] The facial image data FGD are data represented by a structured three-dimensional model of a head of a person HMN wherein components such as a mouth, eyes, a nose and ears, skin, muscle and skeleton can move (See Fig. 8). The persons HMN may be various actual or fictional humans, for example, celebrities such as actors, singers, other artists or stars, sport-players and politicians, ancestors of the user and historical figures. It is also

[0064] The facial image data FGD as described above can be produced by various known methods described below.

[0065] First, three-dimensional shape data are obtained by using any one of following methods. for example.

[0066] (1) A method of presuming a structured facial image based on an ordinary two-dimensional photograph of a face.

[0067] (2) A method of calculating a three-dimensional

shape by using a plurality of two-dimensional images and data indicating a positional relationship between a subject and a camera used for photographing the images (Stereo photography method).

[0068] (3) A method of three-dimensional measurement of a human or a statue by using a three-dimensional measuring apparatus.

[0069] (4) A method of producing a three-dimensional computer graphics character anew.

[0070] Then, the obtained three-dimensional shape data are converted into a structured three-dimensional model. For the conversion, it is possible to employ methods disclosed in Japanese Unexamined Patent Publication No. 8-297751 and Japanese Unexamined Patent Publication No. 11-328440, and a method disclosed in Japanese Patent Application No. 2000-90629 proposed by the present applicant, for example.

[0071] Thus, the structured three-dimensional model is obtained. The form of the structured three-dimensional model can be changed by manipulating its construction points or control points.

[0072] Generally, a skin model is used as a three-dimensional model. Muscle and skeleton may be added to the skin model to generate a three-dimensional model. In the three-dimensional model with the muscle and the skeleton, motion of a person can be expressed more realistically by manipulating the construction points or the control points in the muscle or the skeleton. The data of the three-dimensional model mentioned above are the facial image data FGD. A list LST described below is

prepared with respect to the facial image data FGD accumulated in the facial image database FDB. The each facial image data FGD can be specified by a person number NUM or the like in the list LST.

[0073] As shown in Fig. 5, the list LST is a database for storing information of a plurality of persons HMN who can be persons with whom the user converses. The list LST includes a plurality of fields, for example, the person number NUM for discriminating each of the persons HMN, a person name NAM as a name of the person corresponding to the person number NUM and a sample image SMP indicating an example of a facial image. The list LST stores data concerning the persons HMN such as a person HMN1 and a person HMN2.

[0074] Next, processes and operations performed in the communication system 1 at conversing with a person HMN will be described with reference to flowcharts.

[0075] Fig. 6 is a flowchart showing a process of the communication system 1 of a first embodiment. Fig. 7 is a flowchart showing a process for generating facial animation data FAD and a second message MG2. Fig. 8 generally shows an example of facial image data FGD1.

[0076] As shown in Fig. 6, a user operates a mouse 23a or a keyboard 23b in a client 2 to select from a list LST a person HMN with whom the user converses (#11). A person number HMN of the selected person HMN is transmitted to a server 3 at this point. The list LST may be provided from the server 3 via a network 4, previously stored in a magnetic disk unit 27 as shown in Fig. 1 or provided by media such as a CD-ROM, a floppy disk or the like.

030378207 · 102180

[0077] In the server 3, animation of a person HMN to be displayed before starting a conversation is generated. First, facial image data FGD and person information HMJ corresponding to data of the received person number HMN are extracted from a facial image database FDB and a person information database HDB (#12).

[0078] Next, facial animation data FAD are generated based on the extracted facial image FGD and the person information HMJ (#13) so as to be transmitted to the client 2 (#14). In the client 2, the received facial animation data FAD are displayed on a display 22a as an initial state of the person HMN (#15).

[0079] The second message MG2 may be generated along with the production of the facial animation data FAD so as to be transmitted to the client 2 together with the facial animation data FAD. Further, in the client 2, the second message MG2 may be output from a speaker 22b at the same time with displaying the facial animation data FAD.

[0080] A method for generating the facial animation data FAD and the second message MG2 will be described later in this specification.

[0081] The user watches the person HMN displayed on the display 22a to talk to the person HMN. Specifically, in the client 2, a first message MG1 is input via a microphone 23c or the keyboard 23b so that the input first message MG1 is transmitted to the server 3 (#16).

[0082] The user may start the conversation first, with omitting the steps #13 to #15.

[0083] In the server 3, next facial animation data FAD and the second message MG2 are generated based on the

received first message MG1, the facial image data FGD and the person information HMJ (#17) so that the generated data are transmitted to the client 2 (#18).

[0084] In the client 2, the display 22a or the speaker 22b outputs the facial animation data FAD and the second message MG2 (#19).

[0085] In the case where a disconnection request for stopping the conversation with the person HMN is caused (Yes in #20), the process is finished. On the other hand, if no disconnection request is caused, the process returns to the step #16 so that the conversation (dialogue) between the user and the person HMN is repeated.

[0086] Here, a method for generating the animation or the like performed in the steps #13 and #17 is described.

[0087] The facial image data FGD used in the present embodiment are data represented by a three-dimensional model wherein components such as a mouth, eyes, a nose and ears, skin, muscle and skeleton are structured so as to move.

[0088] The facial image data FGD1 shown in Fig. 8 illustrates a three-dimensional model of skin. The three-dimensional model of skin comprises multiple polygons for forming the skin of the face (head) of the person HMN and a plurality of control points PNT for controlling facial motions.

[0089] Turning to Fig. 7, the received first message MG1 is recognized in the server 3 (#31). In the case where the first message MG1 comprises character data TXT1, it is unnecessary to perform a language recognition process 3gn. If the first message MG1 comprises voice

data SND1, the language recognition process 3gn is performed by using a language recognition conversation engine EG1 so as to generate character data TXTa. If, however, the first message MG1 is not received yet as shown in the step #13, or if the conversation is interrupted for a predetermined period of time, the step #31 is omitted.

[0090] The second message MG2 is generated in order to respond to the first message MG1. Specifically, a conversation generating process 3ki is performed by using the language recognition conversation engine EG1 so as to generate character data TXT2 (#32), and voice data SND2 are then generated based on the produced character data TXT2.

[0091] The character data TXT2 are generated with reference to the character data TXTa or TXT1, sentence information BNJ and word information TNJ. In the case where the character data TXTa or TXT1 are 'How are you?', for example, sentence information BNJ having possibilities that the person HMN responds to the question is extracted from a conversation database KDB with reference to the person information HMJ so as to apply the word information TNJ to the sentence information BNJ. Thus, the character data TXT2 such as 'Fine, thank you. How about yourself?' or 'OK, but I am a little bit tired. Are you all right?' are generated.

[0092] Conversion from the character data TXT2 to the voice data SND2 is performed by using known techniques. However, if the first message MG1 is not received yet as shown in the step #13, or if the conversation is

interrupted for a prejudged period of time, character data TXT2 having possibilities that the person HMN talks to the user are generated with reference to the person information HMJ, the sentence information BNJ, and the word information TNJ in the step #32. Such character data TXT2 include 'Hello.' or 'Is everything OK with you?'.

[0093] Motion control data DSD are produced by using an animation engine EG2 (#34) so as to generate the facial animation data FAD (#35). The motion control data DSD are obtained by executing a motion control process 3ds.

[0094] For example, it is possible to synchronize the facial image data FGD with the voice data SND2 by utilizing the technique disclosed in Japanese Unexamined Patent Publication No. 10-293860 that is described in description of the prior art of the present specification. The facial image data FGD are caused to move based on the motion control data DSD by performing an animation generating process 3an, to thereby generate of the facial animation data FAD.

[0095] In the case of the facial image data FGD1 shown in Fig. 8, the facial image data FGD are caused to move by controlling the control points PNT.

[0096] To send the facial animation data FAD, the data may be compressed by, for example, the MPEG or like encoding methods.

[0097] As described above, according to the first embodiment, facial animation data FAD are generated by a server 3 so as to be transmitted to a client 2. Since the client 2 have only to receive and display the generated data, burden accompanying the data processing is

relatively small. Accordingly, even if the client 2 has difficulties with production of animation due to low performance or low specifications thereof, it is possible to perform a conversation with a person HMN by using the client 2.

(Second Embodiment)

[0098] A whole structure of a communication system 1B of a second embodiment is the same as in the first embodiment, therefore, Fig. 1 is also applied to the second embodiment. However, the second embodiment differs from the first embodiment in a program that is stored in a magnetic disk unit 27 of a client 2B and a magnetic disk unit 37 of a server 3B, and contents processed by processors 21 and 31.

[0099] Specifically, in the first embodiment, facial image data FGD extracted from a facial image database FDB in the server 3B are temporarily stored in a RAM 31B or the magnetic disk unit 37 in the server 3B. In turn, in the second embodiment, the facial image data FGD are transmitted to the client 2B so as to be temporarily stored in a RAM 21b or the magnetic disk unit 27 in the client 2B. Then, facial animation data FAD are generated in the client 2B based on motion control data DSD transmitted from the server 3B.

[0100] Fig. 9 shows an example of a program stored in the magnetic disk unit 27 according to the second embodiment. Fig. 10 shows an example of a program stored in the magnetic disk unit 37 of the second embodiment.

[0101] In Figs. 9 and 10, portions having the same function as in the first embodiment is denoted by the same reference characters and descriptions therefor are omitted

or simplified. The same thing can be applied to other drawings in the present embodiment.

[0102] As shown in Fig. 9, the magnetic disk unit 27 stores an animation generating process 3an for generating animation of a person's face and the facial image data FGD as well as the motion control data DSD transmitted from the server 3B.

[0103] As shown in Fig. 10, a server conversation program 3p stored in the magnetic disk unit 37 comprises a basic operation process 3bs, a language recognition conversation engine EGL and an animation engine EG2 in the same manner as in the first embodiment. Although the animation engine EG2 performs a motion control process 3ds, the animation generating process is not performed therein.

[0104] Next, processes and operations performed in the communication system 1B at conversing with the person HMN will be described with reference to flowcharts.

[0105] Fig. 11 is a flowchart showing a process of the communication system 1B of the second embodiment.

Fig. 12 is a flowchart showing a process for generating the motion control data DSD and a second message MG2.

[0106] As shown in Fig. 11, in the client 2B, a person HMN with whom a user converses is selected from a list LST (#41). A person number NUM of the selected person HMN is sent to the server 3B at this point. After reception of the person number NUM, the server 3B reads facial image data FGD corresponding to the person number NUM from the facial image database FDB so as to transmit the facial image data FGD to the client 2B (#42). Such preprocesses for performing a conversation are automatically carried

out as background processes.

[0107] In the server 3B, the motion control data DSD are generated (#43) so as to be transmitted to the client 2B (#44). In the client 2B, the facial image data FGD are caused to move based on the motion control data DSD, thereby, the facial animation data FAD are generated at the same time with being displayed on a display 22a (#45).

[0108] In addition, the second message MG2 and the motion control data DSD may be concurrently generated in the server 3B so as to be transmitted to the client 2B, and the second message MG2 may be output from a speaker 22b together with display of the facial animation data FAD in the client 2B.

[0109] A first message MG1 is input in the client 2B for transmission to the server 3B (#46). In the server 3B, the motion control data DSD and the second message MG2 are generated based on the first message MG1 and person information HMJ (#47). The generated data are sent to the client 2B (#48).

[0110] The facial image data FGD are output to the display 22a with the data being caused to move based on the motion control data DSD, and at the same time, the second message MG2 is output to the display 22a or the speaker 22b (#49).

[0111] The conversation between the user and the person HMN is repeated until a disconnection request is caused (#46-#50).

[0112] Referring to Fig. 12, a method for generating the motion control data or the like that are performed in the steps #43 and #47 is described.

[0113] In the server 3B, a received first message MG1 is recognized (#61). Character data TXT2 are generated (#62) and voice data SND2 are produced based on the generated character data TXT2 so that the second message MG2 is generated (#63). In addition, the motion control data DSD are generated by using the animation engine EG2.

[0114] As described above, in the communication system 1B of the second embodiment, facial image data FGD extracted at the server 3B are transmitted to the client 2B. Then, in the client 2B, the facial image data FGD are caused to move based on motion control data DSD so that animation is produced. Thus, it is possible to reduce communications traffic of data between the server 3B and the client 2B and to display animation at a high speed according to the second embodiment.

(Third Embodiment)

[0115] A whole structure of a communication system 1C according to a third embodiment is the same as in the second embodiment. Accordingly, Fig. 1 is also applied to the third embodiment. Contents of programs memorized in magnetic disk units 27 and 37 are substantially the same as those of the second embodiment shown in Figs. 9 and 10. However, since data stored in the magnetic disk units 27 and 37 that are provided in a client 2C and a server 3C are different from those of the second embodiment, contents processed by the client 2C and the server 3C are somewhat different.

[0116] More specifically, in the third embodiment, a facial image database FDB is provided in the client 2C and the client 2C performs extraction and temporary storage of

facial image data FGD and generation of facial animation data. The server 3C generates motion control data DSD and a second message MG2 based on a first message MG1 sent from the client 2C.

[0117] Fig. 13 shows an example of databases provided in the magnetic disk unit 27 of the client 2c and the magnetic disk unit 37 of the server 3C according to the third embodiment.

[0118] As shown in Fig. 13, the facial image database FDB is provided only in the magnetic disk unit 27 of the client 2c, and not provided in the magnetic disk unit 37 of the server 3C.

[0119] The process contents in the communication system 1C of the third embodiment are substantially the same as those shown in the flowchart of Fig. 11 of the second embodiment. Only differences will be described below.

[0120] In the step #42 shown in Fig. 11, the facial image data FGD are read from the facial image database FDB provided in the magnetic disk unit 27 of the client 2C for temporary storage. The transmission of the facial image data FGD is not performed. Other process contents in the communication system 1C are the same as those shown in Fig. 11.

[0121] As described above, in the communication system 1C of the third embodiment, the provision of the facial image database FDB in the magnetic disk unit 27 of the client 2C eliminates the need to transmit the facial image data FGD from the server 3C. Accordingly, it is possible to shorten the time taken to start a conversation.

[0122] According to the three embodiments described above,

it is possible to converse remotely with a fictional person or the like with reducing load of processes performed by the client 2 since the second message MG2 is produced in the server 3.

[0123] Since the facial image data FGD are structured in three-dimensional, motion and emotional expressions of the face are variable and natural. Facial animation representing understanding about what a user talks responds to the user with emotional expressions comprised of three-dimensional images and voices and, therefore, the user can enjoy interactive talk.

[0124] In addition, it is possible to realize service of conversing with historical figures and late blood-relative by the selection of the person HMN. When a user selects 'ancestors' from the facial image database FDB as the person HMN, for example, the user can realistically enjoy conversing with the facial animation of the late ancestor.

[0125] In the case where a person HMN is an actual celebrity, a conversation between the celebrity and a plenty of fans can be realized without bothering the celebrity's private life.

[0126] In the language recognition conversation engine EG1, contents of a conversation are set in accordance with kinds of persons HMN such as ancestors, celebrities, historical figures and the like. Thus, a meaningful conversation can be performed between the user and the person HMN.

[0127] Further, by keeping the server 3 in constant operation, the user can enjoy conversing with a person HMN irrespective of time and place.

[0128] Additionally, since the maintenances of the conversation database KDB can be carried out in the server 3, it is possible to easily respond to up-to-date topics, vogue phrases and the like without special maintenances in the client 2.

[0129] In the above-described embodiments, the server 3 generates voice of a person HMN. However, it is also possible to generate only the character data TXT2 at the server 3 and produce the voice data SND2 at the client 2.

[0130] A workstation or a personal computer can be used as the server 3 and the client 2 in the above-described embodiments. As the client 2, there can be used devices with communication facility such as a portable phone, mobile devices and like devices.

[0131] Each part or whole part of structure, circuit, process contents, processing order and contents of a conversation in the communication systems 1, 1B and 1C can be suitably modified in accordance with the spirit and scope of the present invention.

[0132] Other two embodiments of a communication system will be described. In communication systems 1D and 1E according to the two embodiments, a conversation is performed with watching facial animation that is a partner's avatar (substitute) instead of an actual facial image of the partner of conversation. Although a personal computer is used as a terminal device in the communication systems 1D and 1E, other communication equipment such as a telephone, a portable phone, mobile devices can be used as the terminal device.

(Fourth Embodiment)

[0133] Fig. 14 is a block diagram showing a whole structure of a communication system 1D according to a fourth embodiment of the present invention. Fig. 15 shows an example of a program and data stored in terminal devices 5D and 6D of the fourth embodiment. Fig. 16 shows an example of a program and data stored in a host computer 3D of the fourth embodiment.

[0134] As shown in Fig. 14, the communication system 1D comprises the terminal devices 5D and 6D, the host computer 3D and a network 4. A plurality of terminal devices is provided in the communication system 1D and only the terminal devices 5D and 6D are illustrated in Fig. 14.

[0135] The terminal devices 5D and 6D each include a processor 21, a display 22a, a speaker 22b, a mouse 23a, a keyboard 23b, a microphone 23c, a communication controller 24, a CD-ROM drive 25, a floppy disk drive 26 and a magnetic disk unit 27.

[0136] The processor 21 has a CPU 21a, a RAM 21b and a ROM 21c and serves to carry out a series of processes in the terminal devices 5D and 6D.

[0137] The RAM 21b temporarily stores a program, data and the like and the ROM 21c stores a program, information about setting of hardware of the terminal devices 5D and 6D and the like. The CPU 21a executes the programs.

[0138] The display 22a is used for displaying facial animation and the speaker 22b is used for outputting voice of a partner. The mouse 23a and the keyboard 23b are used for operation of the terminal devices 5D and 6D and the

microphone 23c is used for inputting voice.

[0139] The communication controller 24 controls transmission and reception of facial image data FGD as three-dimensional shape data of a face, motion control data DSD used for controlling the facial image data FGD in such a manner that the facial image data FGD move in accordance with a timing of the output of the voice, voice data SND obtained by digital conversion of voice and other data. The CD-ROM drive 25, the floppy disk drive 26 and the magnetic disk unit 27 all stores data and a program.

[0140] The host computer 3D includes a processor 31, a display 32, a mouse 33a, a keyboard 33b, a communication controller 34, a CD-ROM drive 35, a floppy disk drive 36 and a magnetic disk unit 37.

[0141] The processor 31 has a CPU 31a, a RAM 31b, a ROM 31c and the like. The structure and the function of the processor 31 are the same as in the processor 21 described above.

[0142] The network 4 may comprise a public line, a private line, a LAN, a wireless line or the Internet. Each of the terminal devices 5D and 6D is connected to the host computer 3D via the network 4.

[0143] As shown in Fig. 15, each of the the magnetic disk units 27 of the terminal devices 5D and 6D stores an OS 5s as a basic program of the terminal device, a terminal communication program 5p as an application program of the terminal device in the communication system 1D, other necessary programs and data.

[0144] The terminal communication program 5p includes programs such as a basic process program 5pk and a display

process program 5ph or a module. The basic process program 5pk performs processes concerning operations at a user's side such as linkage with the OS 5s, choice of the facial image data FGD and the like. The display process program 5ph serves to move the facial image data FGD based on the motion control data DSD in order to generate animation.

[0145] The programs are suitably loaded into the RAM 21b and executed by the CPU 21a. The received facial image data FGD, the received motion control data DSD and the received voice data SND are stored in the RAM 21b. In addition, the data are stored in the magnetic disk unit 27, if required.

[0146] The display control portion EM1 which is a series of systems for displaying animation is realized as a result of the execution of the various programs on the RAM 21b as described above.

[0147] As shown in Fig. 16, the magnetic disk unit 37 provided in the host computer 3D stores an OS 3s as a basic program of the host computer 3D, a host communication program 3Dp that is an application program of the host computer in the communication system 1D and other necessary programs and data. A facial image database FDB is provided for accumulating the facial image data FGD.

[0148] The host communication program 3Dp includes programs such as a basic process program 3pk, a motion control program 3pd and a language translating program 3py or a module. The basic process program 3pk performs linkage with the OS 3s, supervises and controls an

animation engine EM2 and a language translation engine EM3. The motion control program 3pd generates the motion control data DSD based on the voice data SND. The motion control data DSD are control information used for controlling the facial image data FGD in such a manner that the facial image data FGD move in accordance with a timing of the output of the voice based on the voice data SND. The language translating program 3py is used for translation from voice data SND of a natural language to voice data SND of another natural language.

[0149] The programs are suitably loaded into the RAM 31b and executed by the CPU 31a. Data such as the received voice data SND are stored in the RAM 31b.

[0150] The animation engine EM2 as a series of systems for generating the motion control data DSD and the language translation engine EM3 as a series of systems for translating the voice data SND to another language are realized as a result of the execution of the various programs on the RAM 31b as described above.

[0151] Original voice data are sometimes referred to as 'voice data SND1' and translated voice data are sometimes referred to as 'voice data SND2' in order to be distinguished from each other in the present specification. As to automatic translation of languages, reference may be given to Japanese Unexamined Patent Publication No. 1-211799, for example.

[0152] The facial image data FGD are data represented by a structured three-dimensional model of a head of a human wherein components thereof such as a mouth, eyes, a nose and ears, skin, muscle and skeleton can move. An example

of the facial image data FGD is shown in Fig. 8. The facial image data FGD and the structured three-dimensional model are as described in the first embodiment.

[0153] Partner's avatar is generated based on the facial image data FGD. As the facial image data FGD, it is possible to use actual or fictional objects such as artists, sport players and like celebrities, historical figures, animals and characters in cartoons in addition to a user's face.

[0154] Next, processes and operations performed in the communication system 1D in the case of a conversation between a user of one terminal device 5D and a user of the other terminal device 6D will be described with reference to flowcharts.

[0155] Fig. 17 is a flowchart showing a process of the communication system 1D of the fourth embodiment.

Fig. 18 is a flowchart showing a process of the terminal devices 5D and 6D. Fig. 19 is a flowchart indicating a process of the host computer 3D.

[0156] First, communication between the terminal devices 5D and 6D is established (#110). In order to establish the communication, for example, a request for connection with the terminal device 6D is sent from the terminal device 5D to the host computer 3D. The host computer 3D notifies the terminal device 6D that a connection request is sent from the terminal device 5D. In the case where the connection is permitted, the terminal device 6D performs notification indicating the permission. Various known protocols can also be used for the communication.

[0157] After establishment of the communication, the

host computer 3D transmits partner's facial image data FGD to the terminal devices 5D and 6D as shown in Fig. 17 (#111). Specifically, facial image data FGD selected by the user of the terminal device 6D are sent to the terminal device 5D, and facial image data FGD selected by the user of the terminal device 5D are transmitted to the terminal device 6D. Each of the users selects facial image data FGD according to the user's preference from the facial image database FDB or a database wherein facial image data FGD for each user are previously registered. In the selection, a list of selectable facial image data FGD may be displayed on a display of the each user, or the user may designate facial image data FGD the user like by specifying number or the like. Alternatively, one facial image data FGD previously designated by the each user may be transmitted.

[0158] The users start a conversation (#112). When the conversation is performed, the voice data SND are transmitted from one terminal device to the other terminal device.

[0159] At this point, each of the users can designate a language to be used for speaking and listening with respect to the host computer 3D. If a conversation in English is desired, English is designated as a language to be used for speaking as well as listening. It is also possible to so designate languages that the user can speak in Japanese and listen in English. The user can change the designated language to other languages in the middle of the conversation.

[0160] The host computer 3D judges whether translation is

required in the conversation in accordance with designation of languages sent from the terminal devices 5D and 6D (#113). When a language used by one user for speaking is different from a language used by the other user for listening, the host computer 3D judges that translation is required. In the case where there is no designation of languages, the host computer 3D judges that a specific language, for example, Japanese is used in the conversation.

[0161] In the case where translation is required, the host computer 3D translates by means of the language translation engine EM3 (#114). The voice data SND2 are generated from the voice data SND1 by the translation.

[0162] The motion control data DSD are generated based on the voice data SND (#115). In the case where the translation is performed, the motion control data DSD are generated based on the translated voice data SND2.

[0163] In order to generate the motion control data DSD, for example, information such as phoneme is extracted from the voice data SND for designating words or emotions so that the motion control data DSD are generated by calculating motion of each control point PNT in the facial image data FGD.

[0164] The user may operate the keyboard 23b or the like of the terminal device so as to directly designate the user's emotions, instead of the designation by extracting the emotions from the received voice data SND. In this case, the terminal devices 5D and 6D transmit control data indicating emotions such as 'smile', 'anger' and the like. Thus, even if the user is tired, it is possible to display

0002427282860

animation wherein the user seems to be cheerful on the screen of the receiver.

[0165] The voice data SND and the motion control data DSD are sent from the host computer 3D to the terminal device (#116).

[0166] In the terminal device, the received voice data SND are output from the speaker 22b, and the facial image data FGD that are received first are caused to move based on the received motion control data DSD, thereby producing animation and displaying the animation on the display 22a (#117).

[0167] When the user of the terminal device 5D says 'Good morning', for example, the host computer 3D generates motion control data DSD for giving motion of 'Good morning' to a mouth of the facial image data FGD, and the generated motion control data DSD are transmitted to the terminal device 6D. In the terminal device 6D, a voice of 'Good morning' that is given by the user of the terminal device 5D is output from the speaker 22b. The display 22a displays the facial image data FGD of the user in the terminal device 5D and the mouth thereof opens and closes in connection with a voice of 'Good morning'.

[0168] Additionally, the host computer 3D analyzes emotions of the user in the terminal device 5D based on a tone of 'Good morning'. For example, in the case where the host computer 3D analyzes that the user in the terminal device 5D has a congenial atmosphere, the host computer 3D generates motion control data DSD for moving eyes and a whole face of the facial image data FGD to cause the eyes and the whole face of the facial image data

FGD to smile and then transmits the generated motion control data DSD to the terminal device 6D. Thus, the display 22a in the terminal device 6D displays animation wherein the user in the terminal device 5D says 'Good morning' with smiling.

[0169] As described above, respective users can listen to the partners' voices and watch animation wherein expressions change based on the partners' talks.

[0170] The above-described processes are repeated and the users perform a conversation with watching animations until any one of the users requests disconnection of the communication (#118).

[0171] As shown in Fig. 18, each of the terminal devices 5D and 6D receives the facial image data FGD of a user as a receiver from the host computer 3D (#121). If each of the user starts to talk (Yes in #122), the terminal devices 5D and 6D transmit the voice data SND to the host computer 3D (#123). In the case of receiving the motion control data DSD and the voice data SND (Yes in #124), voice generated thereby is output and animation is displayed (#125).

[0172] As shown in Fig. 19, the host computer 3D sends the facial image data FGD of the user as the receiver to the respective terminal devices 5D and 6D (#131). In the case where the voice data SND are received from the terminal devices 5D and 6D (#132), the host computer 3D carries out translation if required (#133 and #134) and generates the motion control data DSD are generated (#135) followed by transmitting the motion control data DSD and the voice data SND to the respective user's terminal

devices (#136).

[0173] Further, communication can be performed among three or more terminal devices. In this case, facial image data FGD of all other users are transmitted to respective users. Voice of each of the users is transmitted to the terminal devices of the all other users along with motion control data DSD based on the voice. In each of the terminal devices, only animation corresponding to the talking user may be selected from animation based on the received plural facial image data FGD to be displayed. Alternatively, animation of the all users may be simultaneously displayed or may be switched to be displayed one by one.

[0174] According to a communication system 1D of the fourth embodiment, facial image data FGD having a large amount of data are transmitted only once, and only motion control data DSD are sent afterward. Therefore, reduction in communications traffic is realized and it is possible to perform a conversation with watching partner's animation in which a motion appears smooth and substantially natural.

[0175] Since the facial image data FGD are represented by a structured three-dimensional model and three-dimensional animation is displayed on a screen, the display of realistic image close to original image is achieved.

[0176] The provision of an animation engine EG2 in a host computer 3D enables reduction in load of the processes performed by terminal devices 5D and 6D.

[0177] In addition, it is possible to perform a

conversation free from discomfort even in a conversation with a receiver using a different language by providing translation service performed by the host computer 3D.

[0178] Since the motion control data DSD are generated based on the translated voice data SND, translated voice can be satisfactorily coincided with animation.

[0179] For example, motion of a mouth and a face differs by languages. In the case of display of real facial image, motion of a face is not retouched although voice is translated. According to the present embodiment, however, motion of a mouth and a face can be matched with the translated voice. Therefore, it is possible to display natural animation wherein expressions are precisely reproduced along with the translated voice on a screen of a receiver.

[0180] It is also possible to eliminate unnaturalness typically found in dubbed foreign movies, that is caused by discordance of motion of images and voices of different languages or by difference in lengths of voices.

[0181] In the fourth embodiment, original voice data SND1 may be transmitted along with translated voice data SND2. Thereby, multiplexing of voice can be realized so that a user can listen to the translated voice with confirming the original voice.

[0182] Text data of the translated voice data SND2 may be sent along with the translated voice data SND2 so that translated sentences can be displayed along with animation in the terminal devices 5D and 6D.

[0183] In the case where translation is unnecessary, a language translating program 3py and a language

translation engine EM3 in the host computer 3D may be deleted.

(Fifth Embodiment)

[0184] Fig. 20 shows an example of a program and data stored in each of terminal devices 5E and 6E of a fifth embodiment. Fig. 21 shows an example of a program and data stored in a host computer 3E according to the fifth embodiment.

[0185] A whole structure of a communication system of the fifth embodiment is the same as in the fourth embodiment. Differences between the fourth embodiment and the fifth embodiment are programs and data stored in the terminal devices 5E and 6E as well as the host computer 3E and contents processed by processors 21 and 31.

[0186] Specifically, in the fourth embodiment, the facial image database FDB, the motion control program 3pd and the animation engine EM2 are provided in the host computer 3D. In the fifth embodiment, however, a facial image database FDB, a motion control program 5pd and an animation engine EM2 are provided in each of the terminal devices 5E and 6E, as shown in Fig. 20. Therefore, the facial image database FDB, the motion control program 5pd and the animation engine EM2 are not provided in the host computer 3E, as shown in Fig. 21.

[0187] Fig. 22 is a flowchart showing a process of a communication system 1E according to the fifth embodiment. Fig. 23 is a flowchart showing a process of each of the terminal devices 5E and 6E. Fig. 24 is a flowchart showing a process of the host computer 3E.

[0188] As shown in Fig. 22, communication is established

between the terminal devices 5E and 6E, first (#140). Facial image data FGD of respective users of the terminal devices 5E and 6E are exchanged (#141).

[0189] In order to start a conversation, a judgment is made as to whether translation is required (#142). The judgment is performed by the host computer 3D in the fourth embodiment, while the judgment is made by the terminal devices 5E and 6E in the fifth embodiment. For example, in the step #141, each of the users of the terminal devices 5E and 6E sends information of language he/she uses to the other user together with the facial image data FGD. Each of the terminal devices of the users as receivers judges whether translation is required or not based on the received information.

[0190] If it is judged that translation is required after starting a conversation, voice data SND are sent to the host computer 3E (#143). The host computer 3E translates the received voice data SND (#144) and transmits the translated voice data SND to both of the users' terminal devices (#145). If translation is not required, the voice data SND are transmitted to the users' terminal devices (#146).

[0191] Each of the terminal devices generates motion control data DSD based on the received voice data SND (#147). Then, each of the terminal devices outputs the received voice to cause the facial image data FGD to move based on the generated motion control data DSD, thereby displaying generated animation (#148).

[0192] As shown in Fig. 23, each of the users of the terminal devices 5E and 6E can receive the facial image

data FGD (#151) of the user at the other end (a partner) and transmits his/her facial image data FGD to the partner (#152).

[0193] At this time, the received facial image data FGD may be saved in the facial image database FDB in order to be used at conversing with the same partner again.

[0194] In the case where translation is required (Yes in #153), the voice data SND are sent to the host computer 3E (#154). If translation is not required, the voice data SND are transmitted to the partner's terminal device (#155).

[0195] When one of the terminal devices receives the voice data SND sent from the other terminal device or the host computer 3E (Yes in #156), the terminal device generates the motion control data DSD (#157) and output the voice based on the data so that animation is displayed (#158).

[0196] As shown in Fig. 24, when the host computer 3E receives the voice data SND sent from one of the terminal devices (#161), the host computer 3E performs translation (#162) so that the translated voice data SND2 are transmitted to the partner's terminal device (#163).

[0197] According to the fifth embodiment, transmission and reception of the motion control data DSD is not required since motion control data DSD are generated not in a host computer 3E but in terminal devices 5E and 6E. Thus, communications traffic can be further reduced.

[0198] In the fifth embodiment, when translation is not required, voice data SND may be constantly transmitted to a partner's terminal device without making a decision

shown in the step #153. In this case, the transmission may be carried out without using the host computer 3E. Accordingly, a communication system can be constructed by using a simple network not by using the host computer 3E.

[0199] In the fourth and fifth embodiments described above, facial image data FGD are previously obtained in the each terminal device in order to start a conversation, and animation is generated based on the motion control data DSD during the conversation. Thus, communications traffic can be reduced and animation expressing a natural motion can be displayed. Further, even if users uses different languages, it is possible to perform a conversation with watching each partner's animation wherein a motion is smooth and substantially natural.

[0200] As described above, multimedia of animation, voice and character data can be structured by outputting text data corresponding to voice data SND to the users.

[0201] It is possible to modify structure, circuit, process contents, processing order and order of communication of each part or whole part of a terminal device, a host computer or communication systems 1D and 1E can be modified without departing from the spirit and scope of the invention.